Short tandem repeat typing technologies used in human identity testing

John M. Butler

BioTechniques 43:Sii-Sv (October 2007)
doi 10.2144/000112582

Short tandem repeat (STR) typing methods are widely used today for human identity testing applications including forensic DNA analysis. Following multiplex PCR amplification, DNA samples containing the length-variant STR alleles are typically separated by capillary electrophoresis and genotyped by comparison to an allelic ladder supplied with a commercial kit. This article offers a brief perspective on the technologies and issues involved in STR typing.

INTRODUCTION

Short tandem repeats (STRs), which are sometimes referred to as microsatellites or simple sequence repeats (SSRs), are accordion-like stretches of DNA containing core repeat units of between two and seven nucleotides in length that are tandemly repeated from approximately a half dozen to several dozen times (1). Although the human genome contains thousands upon thousands of STR markers, only a small core set of loci have been selected for use in forensic DNA and human identity testing (2). Like using a single, common currency in a financial sense, core loci permit equivalent genetic information to be shared and compared. Commercial kits are now available to generate DNA profiles containing these core STR loci (Table 1). Millions of STR profiles are generated worldwide each year by government, university, and private laboratories performing various forms of human identity testing, including DNA databasing, forensic casework, missing persons/mass disaster victim identification, or parentage testing.

With STR typing, PCR is used to recover information from small amounts of available biological material. The relatively short PCR product sizes of approximately 100–500 bp generated with STR testing are generally compatible with degraded DNA that may be present due to environmental insults on the evidentiary biological material found at a crime scene. PCR amplification of multiple STR loci simultaneously, or multiplexing, is possible with different colored fluorescent dyes and different sized PCR products. Use of multiple loci enables a high power of discrimination in a single test without consuming much DNA (e.g., 1 ng or less of starting material). It is worth noting that these core STR loci occur in between genes in which a high degree of variability is tolerated and are thus not directly responsible for physical traits such as hair color or eye color or genetic diseases.

COMMERCIAL KITS

Commercially available kits, which provide premixed primers and a standard master mixture containing the polymerase, enzyme buffers, and dNTPs, simplify generation of STR profiles and provide results on a uniform set of core STR loci to make it possible for national and international sharing of criminal DNA profiles. Commercial kits are preferred in most forensic laboratories over in-house assays even though the kits are more expensive, as these kits help simplify and standardize procedures and remove the burden of PCR component quality control from the busy end user. In addition, STR kits supply allelic ladders containing common STR alleles that have been previously characterized for the number of repeat units via DNA sequencing. These allelic ladders are used to calibrate PCR product sizes to STR repeat number for genotyping purposes. Figure 1 shows the allelic ladder for the widely used AmpF*l*STR Identifiler kit (Applied Biosystems, Foster City, CA, USA) (3) containing 205 alleles across 16 coamplified loci—15 STRs plus an amelogenin sex-typing assay.

The complete process for STR typing includes sample collection, DNA extraction, DNA quantitation, PCR amplification of multiple STR loci, STR allele separation and sizing, STR typing and profile interpretation, and a report of the statistical significance of a match (if observed). In many casework situations, such as sexual assault evidence, DNA mixtures may result from a combination of the victim and perpetrator's bodily fluids and create a complex and challenging result to interpret.

DNA SEPARATIONS

Following PCR amplification, the overall length of the STR amplicon is measured to determine the number of

Biochemical Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA

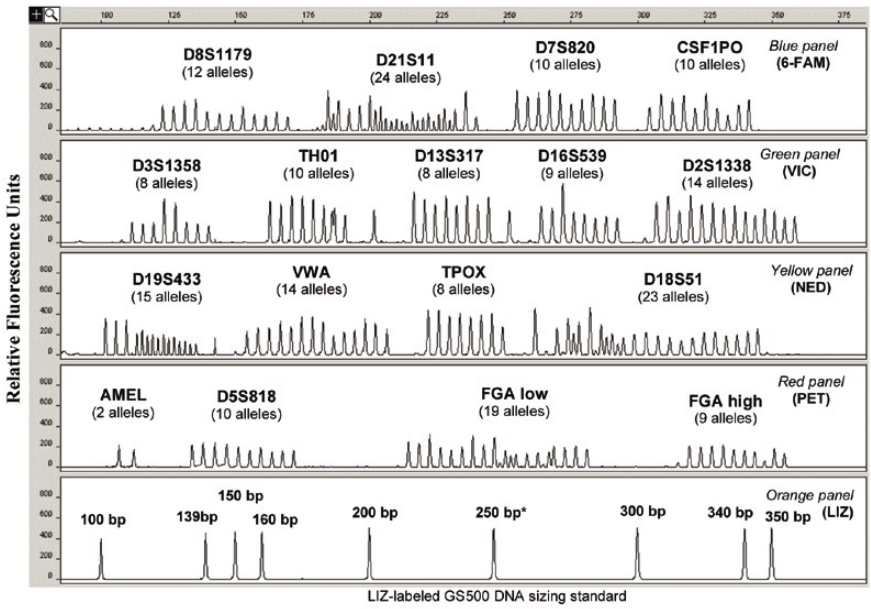

Figure 1. Color separated panels for an allelic ladder from the AmpF/STR Identifiler kit used for DNA size-to-short tandem repeat (STR) calibration. Genotype determination in subsequently processed samples is performed by comparing allele size (relative to an internal size standard) to a commercially provided STR kit allelic ladder with calibrated repeat numbers, which is sized according to the same internal size standard. Note that the 250-bp peak in the GS500 size standard is typically not used due to anomalous migration. Reprinted with permission from Reference 1, Figure 5.6.

repeats present in each allele found in the DNA profile. This length measurement is made via a sized-based separation involving gel or capillary electrophoresis (CE). Each STR amplicon has been fluorescently labeled during PCR, since either the forward or reverse locus-specific primer contains a fluorescent dye. Thus, by recording the dye color and migration time of each DNA fragment relative to an internal size standard, the size for each STR allele may be determined following its separation from other STR alleles. Commonly used instruments for STR allele separation and sizing include the ABI PRISM 310 and ABI PRISM 3100 genetic analyzers (Applied Biosystems) (4).

There are a number of both biological and instrumental artifacts that often must be sorted through in order to generate a complete and accurate STR profile (5; see also Reference 1, Chapters 6 and 15). Biological artifacts include stutter products, split peaks from incomplete adenylation, triallelic patterns, and variant alleles containing

Table 1. Characteristics of the 15 STR Loci Present in the Commercially Available Kit AmpF/STR Identifiler

STR Loci	Chromosomal Location	Repeat Motif	Allele Range[a]	PCR Product Sizes in Identifiler Kit (dye label)
CSF1PO	5q33.1	TAGA	6–15	305–342 bp (6-FAM)
FGA	4q31.3	CTTT	17–51.2	215–355 bp (PET)
TH01	11p15.5	TCAT	4–13.3	163–202 bp (VIC)
TPOX	2p25.3	GAAT	6–13	222–250 bp (NED)
VWA	12p13.31	[TCTG] [TCTA]	11–24	155–207 bp (NED)
D3S1358	3p21.31	[TCTG] [TCTA]	12–19	112–140 bp (VIC)
D5S818	5q23.2	AGAT	7–16	134–172 bp (PET)
D7S820	7q21.11	GATA	6–15	255–291 bp (6-FAM)
D8S1179	8q24.13	[TCTA] [TCTG]	8–19	123–170 bp (6-FAM)
D13S317	13q31.1	TATC	8–15	217–245 bp (VIC)
D16S539	16q24.1	GATA	5–15	252–292 bp (VIC)
D18S51	18q21.33	AGAA	7–27	262–345 bp (NED)
D21S11	21q21.1	[TCTA] [TCTG]	24–38	185–239 bp (6-FAM)
D2S1338	2q35	[TGCC] [TTCC]	15–28	307–359 bp (VIC)
D19S433	19q12	AAGG	9–17.2	102–135 bp (NED)
Amelogenin (sex-typing)	Xp22.22 Yp11.2	Not applicable	Not applicable	X = 107 bp (PET) Y = 113 bp (PET)

The 13 core STR loci used for the U.S. national DNA database are shown in bold font. See www.cstl.nist.gov/biotech/strbase/multiplx.htm for information on other commercially available STR kits.
[a]Ranges are calculated from kit allelic ladders (see Figure 1) and do not represent the full range of alleles observed in world populations. A more complete allele listing of these short tandem repeat (STR) loci is available at www.cstl.nist.gov/biotech/strbase/str_fact.htm.

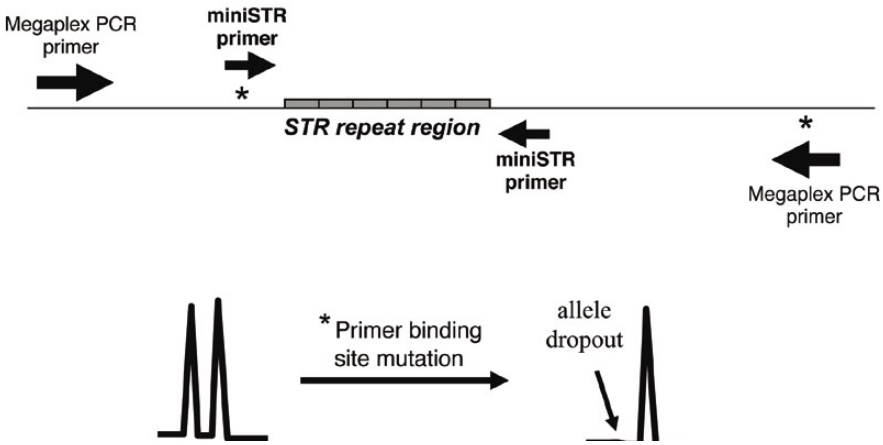

Figure 2. Illustration of PCR primer positions for mini-short tandem repeats (miniSTRs) and megaplex (kit) amplification reactions targeting an STR repeat region (top panel). Generally, the flanking regions around STR repeats are consistent between various alleles, enabling robust amplification of the STR locus. The asterisks represent potential primer binding site mutations that could lead to allele dropout and discordance between the tested primer sets (bottom panel).

mutations in the repeat or flanking regions that cause an allele to be off-ladder. Instrumental artifacts arise from voltage spikes, dye blobs, and bleed-through between dye colors.

While multicolor fluorescence detection CE instrumentation, such as the ABI PRISM 3100 genetic analyzer, presently dominate the field, efforts are ongoing to develop microchip CE platforms (6) to perform high-resolution DNA separations with eventual integration of the PCR amplification and CE separations (7). In addition, mass spectrometry (MS) with matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) techniques have been used for STR typing without allelic ladders (8,9).

miniSTRs: RECOVERY OF INFORMATION FROM DEGRADED DNA

DNA molecules that are exposed to water and/or heat will, over time, begin to break down into smaller pieces. A number of studies have shown what is theoretically predicted—that DNA types can be recovered more effectively from degraded DNA samples when the PCR products are smaller (10–12). By moving the PCR primers closer to the STR repeat region, the product sizes can be reduced while retaining the same information (11). The utility of miniSTR assays has been confirmed in intra- and interlaboratory studies involving degraded bone samples and aged blood and saliva stains (12). Success rates in recovering information from compromised DNA samples improve with miniSTR systems compared with conventional STR kits. A timeline covering the development of miniSTRs may be found at www.cstl.nist.gov/biotech/strbase/miniSTR/timeline.htm.

Using their 5-dye chemistry and mobility modifier technology, Applied Biosystems has developed a miniSTR kit capable of amplifying eight core STR loci and amelogenin with reduced PCR product sizes relative to current commercial kits. This AmpF*l*STR MiniFiler kit, which includes an improved PCR master mixture, should aid efforts to recover results from degraded DNA samples. However, it is important to keep in mind that because different PCR primers are in use with the miniSTR kit relative to previous STR kits, discordant results may occur due to primer binding site mutations that cause allele dropout.

CONCORDANCE TESTING

Some STR alleles contain sequence variation in the flanking regions. If a nucleotide change (or insertion or deletion) occurs in a PCR primer binding site in a particular allele, then it is possible for the mutant sequence to be incompatible with primer annealing and the variant allele will fail to amplify (1). This allelic dropout is sometimes referred to as a null allele. A sample that is really a heterozygote might show up as an apparent homozygote if the PCR primers fail to anneal and amplify the allele containing the point mutation (Figure 2). In some cases, the null allele may be recovered by lowering the annealing temperature, permitting less stringent binding between the primer and DNA template. A concordance study, which tests a set of the same DNA samples with nonoverlapping PCR primers, permits detection of null alleles. An examination of MiniFiler versus Identifiler found only 27 discrepancies in over 10,000 genotypes compared (13). A summary of discordant results observed in concordance studies is available at www.cstl.nist.gov/biotech/strbase/NullAlleles.htm.

As long as the same PCR primers are used, identical STR typing results can be expected on the same DNA sample. However, if primer positions are changed, then the possibility exists for allele dropout or size shifting to occur. In some cases, extra primers matching the mutant allele—often referred to as degenerate primers—can be including in the kit primer mixture to enable recovery of PCR amplification when the variant allele containing a mismatch in the primer binding site is present (14).

PERSPECTIVES ON THE FUTURE

As noted almost 7 years ago by the Research and Development Working Group of the National Commission on the Future of DNA Evidence (15), STRs will probably be the markers of choice for the foreseeable future because of their widespread use in national DNA databases (2,16,17). We will likely see evolutionary progression in typing technology, more rapid amplification methods, additional STR kits containing new loci, and computer programs that will enable fast evaluation of STR typing data. Through continued advances, STR typing technologies may become miniaturized

and integrated with other parts of the process, such as DNA extraction and amplification. Effective DNA databases are being constructed and numerous forensic cases solved today through generating STR profiles with a common set of genetic markers. These current core loci have played and will continue to play a vital role in human identity testing.

ACKNOWLEDGMENTS

The author, John M. Butler, is funded by the National Institute of Justice through interagency agreement 2003-IJ-R-029 with the National Institute of Standards and Technology (NIST) Office of Law Enforcement Standards. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice. Certain commercial equipment, instruments, and materials are identified to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the NIST nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. Comments on this manuscript by Peter Vallone (NIST Human Identity Project Team) were greatly appreciated.

COMPETING INTERESTS STATEMENT

The author declares no competing interests.

REFERENCES

1. **Butler, J.M**. 2005. Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. 2nd ed. Elsevier Academic Press, New York.
2. **Butler, J.M.** 2006. Genetics and genomics of core STR loci used in human identity testing. J. Forensic Sci. *51*:253-265.
3. **Collins, P.J., L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder, and P.A. Foxall.** 2004. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR Identifiler PCR Amplification Kit. J. Forensic Sci. *49*:1265-1277.
4. **Butler, J.M., E. Buel, F. Crivellente, and B.R. McCord.** 2004. Forensic DNA typing by capillary electrophoresis: using the ABI Prism 310 and 3100 Genetic Analyzers for STR analysis. Electrophoresis *25*:1397-1412.
5. **Scientific Working Group on DNA Analysis Methods (SWGDAM).** 2000. Short tandem repeat (STR) interpretation guidelines. Forensic Sci. Comm. 2(3); online at www.fbi.gov/hq/lab/fsc/backissu/july2000/strig.htm.
6. **Yeung, S.H., S.A. Greenspoon, A. McGuckian, C.A. Crouse, C.A. Emrich, J. Ban, and R.A. Mathies.** 2006. Rapid and high-throughput forensic short tandem repeat typing using a 96-lane microfabricated capillary array electrophoresis microdevice. J. Forensic Sci. *51*:740-747.
7. **Liu, P., T.S. Seo, N. Beyor, K.J. Shin, J.R. Scherer, and R.A. Mathies.** 2007. Integrated portable polymerase chain reaction-capillary electrophoresis microsystem for rapid forensic short tandem repeat typing. Anal. Chem. *79*:1881-1889.
8. **Butler, J.M., J. Li, T.A. Shaler, J.A. Monforte, and C.H. Becker.** 1998. Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. Int. J. Legal Med. *112*:45-49.
9. **Oberacher, H., F. Pitterl, G. Huber, H. Niederstatter, M. Steinlechner, and W. Parson.** Increased forensic efficiency of DNA fingerprints through simultaneous resolution of length and nucleotide variability by high-performance mass spectrometry. Hum. Mutat. (In press).
10. **Wiegand, P. and M. Kleiber.** 2001. Less is more—length reduction of STR amplicons using redesigned primers. Int. J. Legal Med. *114*:285-287.
11. **Butler, J.M., Y. Shen, and B.R. McCord.** 2003. The development of reduced size STR amplicons as tools for analysis of degraded DNA. J. Forensic Sci. *48*:1054-1064.
12. **Dixon, L.A., A.E. Dobbins, H.K. Pulker, J.M. Butler, P.M. Vallone, M.D. Coble, W. Parson, B. Berger, et al.** 2006. Analysis of artificially degraded DNA using STRs and SNPs—results of a collaborative European (EDNAP) exercise. Forensic Sci. Int. *164*:33-44.
13. **Hill, C.R., M.C. Kline, J.J. Mulero, R.E. Lagace, C.W. Chang, L.K. Hennessy, and J.M. Butler.** 2007. Concordance study between the AmpFlSTR MiniFiler PCR amplification kit and conventional STR typing kits. J. Forensic Sci. *52*:870-873.
14. **Leibelt, C., B. Budowle, P. Collins, Y. Daoudi, T. Moretti, G. Nunn, D. Reeder, and R. Roby.** 2003. Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles. Forensic Sci. Int. *133*:220-227.
15. **National Institute of Justice (NIJ).** 2000. The future of forensic DNA testing: predictions of the Research and Development Working Group of the National Commission on the Future of DNA Evidence. Washington, D.C. www.ojp.usdoj.gov/nij/pubs-sum/183697.htm.
16. **Gill, P.** 2002. Role of short tandem repeat DNA in forensic casework in the UK—past, present, and future perspectives. BioTechniques *32*:366-372.
17. **Gill, P., D.J. Werrett, B. Budowle, and R. Guerrieri.** 2004. An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci. Justice *44*:51-53.

Address correspondence to John M. Butler, Biochemical Science Division, National Institute of Standards and Technology, 100 Bureau Drive, Mail Stop 8311, Gaithersburg, MD 20899-8311, USA. e-mail: john.butler@nist.gov